# CHAPTER 1

# From Groups to Leaders and Back

## Exploring Mutual Predictability Between Social Groups and Their Leaders

Francesco Solera[,*], Simone Calderara[*] and Rita Cucchiara[*]
[*] University of Modena and Reggio Emilia, Department of Engineering, Italy
[a] Corresponding: francesco.solera@unimore.it

### Abstract

Recently, social theories and empirical observations identified small groups and leaders as *the basic elements which shape the crowd*. This leads to an intermediate level of abstraction that is placed between the crowd as a flow of people, and the crowd as a collection of individuals. Consequently, automatic analysis of crowd in computer vision is also experiencing a shift in focus from individuals to groups and from small groups to their leaders. In this chapter, we present state of the art solutions to the groups and leaders detection problem, which are able to account for physical factors as well as for sociological evidence observed over short time windows. The presented algorithms are framed as structured learning problems over the set of individual trajectories. However, the way trajectories are exploited to predict the structure of the crowd is not fixed but rather learnt from recorded and annotated data, enabling the method to adapt these concepts to different scenarios, densities, cultures and other unobservable complexities. Additionally, we investigate the relation between leaders and their groups and propose the first attempt to exploit leadership as prior knowledge for group detection.

**Keywords:** group detection, leader identification, crowd analysis, structured learning, social computer vision

### Chapter points

- Survey of social theories of crowd, groups and leadership and respective computational approaches.
- Structured learning framework for automatic visual detection of groups and leaders in crowd.
- Empirical experiments to delve deeper into the analysis of the mutual influence between groups and their leaders.

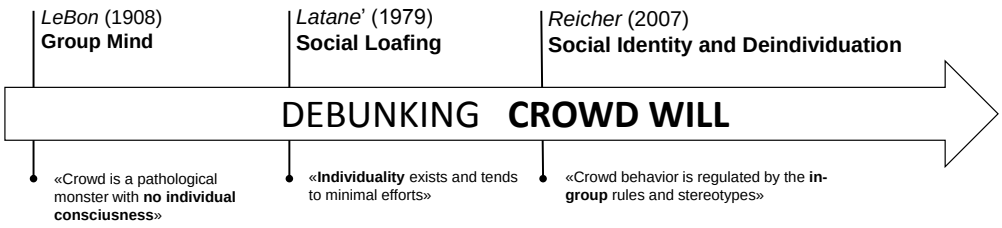
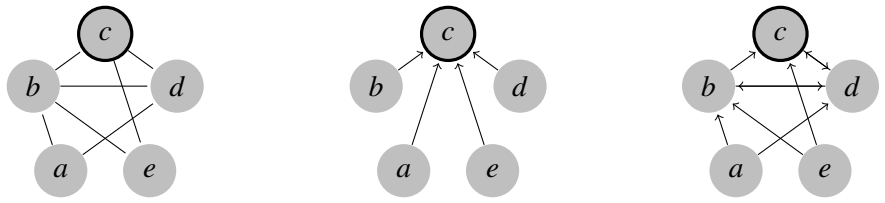



Figure 1.1: Evolution of crowd theories from mass phenomena to groups formation.

## 1. Introduction

Understanding crowd dynamics has engaged many scientists in the past century from different heterogeneous points of view, ranging from collective psychology to system theory, involving sociology and computer vision as the main analytic tools. Crowd phenomena are complex and alluring modern elephant men (Reicher [1]), because their logic still escapes formal rules and contemporarily exposes fascinating challenges. Eventually the ambition is always to precisely characterize people behavior in crowd, to predict and prevent potentially dangerous situations by means of either synthetic simulation models or real time visual analysis. In his pioneering work on crowd behavior, Gustave Le Bon [2] defined crowds as hidden and inherent threats to society. In his writings Le Bon asserted that as members of a crowd, people tend to display a loss of self-awareness and an increased inclination towards violence. Far from this approach, the modern elaborated *Social Identity Model* [1] proposes a social-normative conception of collective behavior based on members spontaneous transition from an individual identity to a common and shared one among small subset of people, also known as *groups*.

In accordance with recent theories, empirical observations [3] recognize groups as the basic elements which the crowd is composed of, leading to an intermediate level of abstraction that is placed between the crowd as a flow of people and its interpretation as a collection of individuals, Figure 1.1. Identifying groups is consequently a mandatory step in order to grasp the complex social dynamics ruling collective behaviors in crowds. Nevertheless, automatic group detection in video streams is definitely less studied than pedestrian analysis or crowd flow motion estimation. One of the greater challenges resides in the lack of a single, agreed, computational definition of a group, formal definitions of the mechanisms which govern them and insights on the relations arising among people during social gatherings. Conversely, there seems to be agreement on the fact that not all the members of a group (and of a crowd, more generally) undergo the same level of identity shift [2, 4, 5]. People who define the norms and the values which then become shared among all the other members are recognized as




(a) Le Bon's theory (b) Freud's theory (c) Tarde's theory

Figure 1.2: Different interpretation of relationships between a leader and other members in different crowd psychology theories.

*leaders*, thus identifying leaders is crucial in crowd management, emergency planning and sociological analysis.
The purpose of this chapter is twofold:

- We aim to provide a learning framework that can be useful for the visual recognition of both groups and leaders. These problems are neatly casted into the same structured learning framework [6, 7], and solved efficiently using Structural SVMs, Sec. 3 and followings.
- We aim to discuss, starting from empirical evidence and original experiments, the roles of leaders in forming and structuring a group and the mutual influence groups and leaders have in their automatic visual detection, Sec. 6.

We hope such a discussion will bring benefit to the computer vision community by raising awareness of the social mechanisms underpinning crowds and groups, and by providing straightforward solutions to leaders and groups detection.

## 2. Modeling and Observing Groups and Their Leaders in Literature

In this section we briefly survey past and recent sociological theories that have tried to explain crowds' behavior and their structure, as well as computational models employed by the computer vision community to automatically analyse crowd and related events.

### 2.1. Sociological Perspective

Most of the research work has tried to tackle the crowd as an exclusively collective phenomenon, where individuality and social groups do not exist. This recalls the primitive *Popular Mind Theory* [2], where the crowd was defined as a "pathological monster with no individual consciousness". Accordingly, crowds have been modeled by means of physical models (*e.g.* hydrodynamics [8]), neglecting the existence of sin-

gle individual purposes and goals. Conversely, many other studies have been inspired by the '70s *Social Loafing Theory* [9], which stated that individuality was a strong requirement for the pursuit of personal goals. Helbing's *Social Force Model* [10], which asserts that anyone movements towards her goals are influenced by the surrounding pedestrians, has also been the main building block for many crowd modeling and analysis works. Recently, studies on pedestrians attending events have underlined that most of the people tend to move in groups and social relations influence the way people behave in crowds [11, 3], Figure 1.1. These empirical observations are supported by Reicher in the recent *Social Identity Model of Deindividuation Effects* [12], which assumes that crowd behavior is regulated by the social rules and behaviors groups choose to adopt. Actual field observations of temporal gatherings by McPhail [13] indicate that members are rarely violent, leaders provide direction through verbal and reasonable interventions, and composing groups do not act in a capricious, unpredictable fashion. These groups form, change, and disband, and the internal structures and processes of the crowd and its groups are more similar than different.

If crowds can be understood by taking into account group processes, then their leaders' actions can be understood by taking into account leadership processes. Le Bon's leadership model [2], resembles the need of the crowd to place trust in someone able to provide orientation and contribute to its overall stability. His leader is neither a founding figure nor he is permanently established, see Figure 1.2(a). Instead, the crowd formation is an emergent process that uses the leader as a stabilizer. While for Le Bon the leader is an elected anonymous figure, in Freud's *Group Psychology* [4] the leader comes to play a constitutive role and every member of the crowd identifies the leader as their "I"-ideal. With respect to Le Bon self-organizing and emergent notion of crowds, Freud provides a highly centralized model of the social community, appreciable in Figure 1.2(b). The relation between the leader and the crowd is thus radically asymmetric as the members are submitted under the leader; relations between crowd members are secondary. Eventually, and to even a greater extent than Le Bon, Tarde [5] emphasizes the self-referential emergence of crowd phenomena but characterizing the leader as the "spark" behind the organizational patterns. Unlike Freud, Tarde was not interested in the leader's foundational role. Instead, he analyzes how the leader contributes to the flow of imitation in a society and builds its theory assuming that every process of imitation begins with asymmetry, as highlighted in Figure 1.2(c). Nevertheless, as individuals initiative converge with leadership, the leader's identity may change without altering the crowd and groups stability [13].

Due to lack of accepted theories on leader emergence in groups and following McPahil's on-the-field observations that crowd and groups share similar and hierarchical formation and structuring processes, in the rest of this work we will borrow sociological theories about crowd leaders and apply them to group leaders.

## 2.2. Computational Approaches

The computational modeling of pedestrian dynamics in crowd situations represents a relatively recent but already established field of research with many different application. Most models employ basic proxemic notions [14], like the tendency of pedestrians to preserve a personal space around them whenever possible. However, recent approaches take a more comprehensive viewpoint on proxemics, that also comprises the preference of an individual to stay close to other members that belong to the same group [15, 16]. Sociological concepts such as *F-formations* by Kendon [17] have been exploited as a foundation for an interesting line of research on group detection [18, 19]. F-formations can be interpreted as specific positional and orientational patterns that people assume in order to be considered engaged in a social interaction. Nevertheless, the theory holds only for stationary groups and is not defined for moving groups, a case which cannot be ignored in crowd analysis. Differently, motion paths are considered as the main feature by the most recent group detection approaches. There are three approaches: In *group-based* approaches, groups are considered as atomic entities in the scene and no higher level information can be extracted neatly, typically due to high noise or high complexity of crowded scenes [20, 21, 22]. Up to now, these models was confined to the detection problem and were not used to further infer on groups paths and identities in time compared to the *individual-group joint* approaches that combined the individual tracking while tracking groups at a coarser level [23, 24]. Finally, *individual-based* approaches build up on single pedestrians trajectories. This kind of approach is subjected to the challenge of tracking individuals even in high density crowds and can be applied with significant limitations in real life scenarios [25, 26, 27, 28].

While visual group detection has gained momentum in the computer vision community, leader identification is still an emerging topic. Pioneering works tackled the problem of leadership identification as finding the group member that contribute the most to the group proxemic formation. This line of work builds on the intuition that leaders cover the central position inside the group [29, 30]. Despite being effective there is still a lack of empirical evidence that the leader spatial centrality holds independently from the crowd type and its density and, as a matter of fact, this is typically neglected in sociological literature. More complex approaches adhered to Tarde's referential model of leadership exploiting either Bayesian inference on causal graph or ranking techniques in the feature space to establish the leader inside a group, [31, 32, 33]. Referential models are up to now the most effective models of leadership. Unluckily, most of the models solve the problem considering one group at the time, and not exploring the mutual relations between members and leaders of different groups. Conversely, in crowd psychology these inter-group relations have been supposed in recent theories.

## 3. Technical Preliminaries and Structured Output Prediction

Many inference tasks related to crowd analysis are structured prediction tasks. Structured output prediction is the inference task related to a set of random variables whose outcome is interdependent in complex but observable ways. Canonical examples of such structured objects may include matrices, sequences or graphs, with applications to image segmentation, natural language processing or bio-informatics, among others. Structured prediction neatly applies to crowds because individual behavior is oftentimes interlinked with the behavior of other people as well[1].

This section introduces Structural Support Vector Machines (SSVM) [34], a discriminative method for complex and structured output prediction. On top of the SSVM learning framework, in Sec. 4 we develop algorithms to detect groups and their leaders in social gatherings and crowded environments.

### 3.1. Problem Statement

Let us consider the input $\mathbf{x} \in \mathcal{X}$ to be some representation of the crowd, possibly describing all unary, pairwise and higher order interaction terms. As an example, imagine $\mathbf{x}$ could describe crowd members' position, mutual distance, engagement and so on. We want to learn a mapping from input $\mathbf{x}$ to output variable $\mathbf{y} \in \mathcal{Y}(\mathbf{x})$ - possibly describing each member social group (Sec. 4.1) or its role (Sec. 4.2) - based on a set of samples $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1\dots n}$ drawn from a fixed but unknown distribution. Depending on the task and on the input $\mathbf{x}$, the output space $\mathcal{Y}(\mathbf{x})$ will have different sizes and characterization.

A discriminant score function $F : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is defined over the joint input-output space, such that $F(\mathbf{x}, \mathbf{y})$ can be interpreted as measuring the compatibility of output proposal $\mathbf{y}$ given a specific input $\mathbf{x}$. Now, the prediction function $f : \mathcal{X} \to \mathcal{Y}$ can be defined as:

$$f(\mathbf{x}) \;=\; \arg\max_{\mathbf{y}\in\mathcal{Y}(\mathbf{x})} F(\mathbf{x}, \mathbf{y}) \tag{1.1}$$

where the maximizer over the label space $\mathcal{Y}(\mathbf{x})$ is the predicted label, *i.e.* the solution of the inference problem. For simplicity we choose to restrict the space of $F$ to linear functions over some combined feature representation $\Psi(\mathbf{x}, \mathbf{y})$ subject to a $\mathbf{w}$-parametrization, so that $F(\mathbf{x}, \mathbf{y}) = \mathbf{w}^T \Psi(\mathbf{x}, \mathbf{y})$.

The problem of learning in structured and interdependent output spaces can been formulated as a maximum-margin problem. We adopt the $n$-slack, margin-rescaling

[1] Here and in the rest of this work we will use the word *behavior* without implying any further sociological claim. Eventually, to what extent behavior is really captured is up to the features extracted from the videos and employed in the presented methods.

formulation:

$$\begin{aligned} \min_{\mathbf{w},\xi} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}\xi_i \\ \text{s.t.} \quad & \forall i : \xi_i \geq 0, \\ & \forall i, \forall \mathbf{y} \in \mathcal{Y}(\mathbf{x}_i)\backslash\mathbf{y}_i : \mathbf{w}^T\delta\Psi_i(\mathbf{y}) \geq \Delta(\mathbf{y},\mathbf{y}_i) - \xi_i, \end{aligned} \tag{1.2}$$

where $\delta\Psi_i(\mathbf{y}) \stackrel{\text{def}}{=} \Psi(\mathbf{x}_i,\mathbf{y}_i) - \Psi(\mathbf{x}_i,\mathbf{y})$, $\xi_i$ are the slack variables introduced in order to accommodate for margin violations, $\Delta(\mathbf{y}_i,\mathbf{y})$ is the loss function measuring distance between two outputs and $C$ is the regularization trade-off. Intuitively, we want to maximize the margin and jointly guarantee that for a given input, every possible output result is considered worst than the correct one by at least a margin of $\Delta(\mathbf{y}_i,\mathbf{y}) - \xi_i$, where $\Delta(\mathbf{y}_i,\mathbf{y})$ is larger when the two predictions are known to be more different.

Note that both the feature map $\Psi(\mathbf{x},\mathbf{y})$ and the loss function cannot be defined out of the context of the problem, as it is the problem itself that specifies (i) given a particular input, the nature of the desired solution; and (ii) how to account for differences in output objects. As a result, SSVM is more of a framework than an off-the-shelf algorithm. Sec. 4 will introduce the feature map and the loss function for the tasks of group detection and leader identification.

## 3.2. Stochastic Optimization

The quadratic program (QP) (1.2) introduces a constraint for every possible wrong prediction of the $n$ examples in the training set $\mathcal{D}$, more precisely $\sum_{i=1}^{n}(|\mathcal{Y}(\mathbf{x}_i)| - 1)$. Unfortunately, the number of possible solutions typically involved in combinatorial objects, such as graphs, scales exponentially (or worse) with the size of the input, making the optimization intractable.

In order to deal with this high number of constraints many approximation schemes have been proposed, where cutting plane algorithms or subgradient methods (*e.g.* [35, 36]) are among the most common. In particular, if for each example we rearrange all the constraints of QP (1.2) and focus on satisfying just the one requiring the highest $\xi_i$, we can define the structured hinge-loss as the highest classification penalty for a specific example:

$$\widetilde{H}(\mathbf{x}_i) \stackrel{\text{def}}{=} \max_{\mathbf{y}\in\mathcal{Y}} \Delta(\mathbf{y}_i,\mathbf{y}) - \mathbf{w}^T\delta\Psi_i(\mathbf{y}). \tag{1.3}$$

The computation of the structured hinge-loss for each element $i$ of the training set amounts to finding the most "violating" output $\mathbf{y}^*$ for a given input $\mathbf{x}_i$ and its correct associated output $\mathbf{y}_i$. Eq. (1.3) suggests the violation resides in having simultaneously a high loss and a high compatibility, which is a contradiction by definition. We are

now left with an unconstrained, non-smooth version of QP (1.2):

$$\min_{\mathbf{w}} \quad \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n} \max\{0, \widetilde{H}(\mathbf{x}_i)\}. \tag{1.4}$$

By disposing of a maximization oracle, *i.e.* a solver for Eq. (1.3), and a computed solution $\mathbf{y}^*$, subgradient methods can easily be applied to QP (1.4), being $\partial_{\mathbf{w}}\widetilde{H}(\mathbf{x}_i) = -\delta\Psi_i(\mathbf{y}^*)$.

To exploit the domain separability of the constraints and limit the number of oracle calls needed to converge to the optimal solution, we choose to adopt a Block-Coordinate version of the Frank-Wolfe algorithm (BCFW) [37], delineated in Alg. 1.

**Algorithm 1** Block-Coordinate Frank-Wolfe Algorithm

1: Let $\mathbf{w}^{(0)}, \mathbf{w}_i^{(0)} := \mathbf{0}$ and $l^{(0)}, l_i^{(0)} := 0$
2: **for** $k \leftarrow 0$ **to** $n_{\text{it}}$ **do**
3: Pick $i$ at random in $\{1, \ldots, n\}$
4: Solve $\mathbf{y}^* \leftarrow \arg\max_{\mathbf{y}\in\mathcal{Y}} \Delta(\mathbf{y}_i, \mathbf{y}) - \mathbf{w}^T\delta\Psi_i(\mathbf{y})$
5: Let $\mathbf{w}_s \leftarrow \frac{C}{n}\delta\Psi_i(\mathbf{y}^*)$ and $l_s := \frac{C}{n}\Delta(\mathbf{y}_i, \mathbf{y}^*)$
6: Let $\gamma \leftarrow \frac{(\mathbf{w}_i^{(\mathrm{k})} - \mathbf{w}_s)^T\mathbf{w}^{(\mathrm{k})} + \frac{C}{n}(l_s - l_i^{(\mathrm{k})})}{\|\mathbf{w}_i^{(\mathrm{k})} - \mathbf{w}_s\|^2}$ and clip to $[0, 1]$
7: Update $\mathbf{w}_i^{(\mathrm{k}+1)} \leftarrow (1-\gamma)\mathbf{w}_i^{(\mathrm{k})} + \gamma\mathbf{w}_s$ and $l_i^{(\mathrm{k}+1)} := (1-\gamma)l_i^{(\mathrm{k})} + \gamma l_s$
8: Update $\mathbf{w}^{(\mathrm{k}+1)} \leftarrow \mathbf{w}^{(\mathrm{k})} + \mathbf{w}_i^{(\mathrm{k}+1)} - \mathbf{w}_i^{(\mathrm{k})}$ and $l^{(\mathrm{k}+1)} := l^{(\mathrm{k})} + l_i^{(\mathrm{k}+1)} - l_i^{(\mathrm{k})}$
9: **end for**

The algorithm works by minimizing the objective function of Eq. (1.4) but restricted to a single random example at each iteration. By calling the max oracle upon the selected training sample (line 4) we obtain a new sub-optimal parameter set $\mathbf{w}_s$ by simple derivation (line 5). The best update is then found through a closed-form line search (line 6), greatly reducing convergence time compared to other subgradient or cutting plane methods.

## 4. The Tools of the Trade in Social and Structured Crowd Analysis

In this section we report and summarize authors' approaches to group detection [6] and leader identification [7] in crowds. Previous work has shown that the concept of *group* and *leader* cannot be uniquely specified, but varies according to the crowdness, the environment, the cultural habits and other factors that are difficult to detect and encode. As a consequence, learning seems an adequate paradigm to tackle social related computer vision tasks. At the same time, the methods presented in the remainder of this section have demonstrated good generalization ability, implying they can be ap-

plied off-the-shelf to scenarios similar to the ones used for training; otherwise a short training stage might be required.

Learning is accomplished through Structural SVM, introduced in Sec. 3. In Secs. 4.1 and 4.2, we delineate the feature map and loss function for the tasks of group detection and leader identification. Eventually, these methods will be the two key ingredients employed in the investigation on the relationships between a leader behavior and the behavior of the rest of the group.

## 4.1. Socially Constrained Structural Learning for Groups Detection in Crowd

As previously described, modern crowd theories agree that collective behavior is the result of the underlying interactions among small groups of individuals and individuals (even if singletons are rarer in some kind of crowds than others). This is why detecting groups of socially connected pedestrians (social groups) already is a central topic in computer vision aided crowd analysis. Here we propose a solution for visually detecting groups in low/medium density crowds under the hypothesis that the *groups* can be visually discerned and people walking paths can be tracked up to some extent. In order to design a computational model, we rely on Turner's definition of groups as *two or more people interacting to reach a common goal and perceiving a shared membership, based on both physical and social identity* [1], and on a set of extracted features grounded on this definition.

### *4.1.1. Task Formulation*

We cast the group detection task as a clustering problem. Consider a set of pedestrian $\mathcal{M} = \{a, b, \dots\}$ and $\mathcal{Y}(\mathcal{M})$ as the set of all possible ways to partition $\mathcal{M}$. Defining $y$ as a subset of pedestrians (also referred to as group or cluster) in $\mathcal{M}$, a generic set of subsets $\mathbf{y} = \{y_1, y_2, \dots\}$ is a valid solution in $\mathcal{Y}(\mathcal{M})$ if the partitioning axioms are satisfied: $\forall a \in \mathcal{M}, \exists! y \in \mathcal{Y}(\mathcal{M}) : a \in y$ and $\cup_{y \in \mathcal{Y}(\mathcal{M})} y = \mathcal{M}$. There are two trivial partitioning solutions: one where all the pedestrians are clustered as a single group ($|\mathbf{y}| = 1$) and one where each pedestrian belongs to different clusters ($|\mathbf{y}| = |\mathcal{M}|$). In the remained of this section we call *singletons* those pedestrians whose cluster is composed by themselves only, *i.e.* $|y| = 1$.

We propose to solve the crowd partitioning problem by employing the *Correlation Clustering* (CC) [38]. The CC algorithm takes as input an affinity matrix $W$ where, if $W^{ab} > 0$ ($W^{ab} < 0$), elements $a$ and $b$ belong to the same (different) cluster with certainty $|W^{ab}|$. The algorithm returns the partition $\mathbf{y}$ of a set of elements $\mathcal{M} = \{a, b, \dots\}$ so that the sum of the affinities between item pairs in the same clusters $y$ is maximized:

$$\mathrm{CC} = \arg \max_{\mathbf{y} \in \mathcal{Y}(\mathcal{M})} \sum_{y \in \mathbf{y}} \sum_{a \neq b \in y} W^{ab}_{\mathbf{d}}. \tag{1.5}$$

The pairwise elements affinity in $W$ is $\mathbf{w}$-parametrized as a weighted linear combination of a bounded dissimilarity measure and its complement:

$$\begin{aligned} W_{\mathbf{d}}^{ab} &= \boldsymbol{\alpha}^T(\mathbf{1} - \mathbf{d}(a,b)) - \boldsymbol{\beta}^T\mathbf{d}(a,b) \\ &= \underbrace{-(\boldsymbol{\alpha}+\boldsymbol{\beta})^T\mathbf{d}(a,b)}_{\text{from distance to correlation}} + \underbrace{\boldsymbol{\alpha}^T\mathbf{1}}_{\text{threshold}} . \end{aligned} \tag{1.6}$$

To be consistent with Turner's definition of groups, we design the pairwise distance between pedestrian $a$ and $b$, $\mathbf{d}(a,b)$ as a vector of pairwise distances built upon different aspects that concur to unveil the presence of groups. In detail, we measure the physical relation between pedestrian $d_{ph}$, their mutual influences in motion pattern by causality and trajectories shape analysis, $d_{ca}$ and $d_{sh}$, and their simultaneous convergence to peculiar zones in the scene $d_{he}$ obtaining $\mathbf{d}(a,b) = [d_{\text{ph}}, d_{\text{sh}}, d_{\text{ca}}, d_{\text{he}}]$. For a deeper presentation and discussion about the employed features, please refer to author's previous work [6].

#### 4.1.2. SSVM Adaptation to Group Detection

By tuning $\mathbf{w} = [\boldsymbol{\alpha}, \boldsymbol{\beta}]$ parameters in Eq. (1.6) we can evaluate many different groupings. In order to efficiently learn those parameters according to different peculiarities groups exhibit in different scenarios, we now introduce the feature map and a loss function specifically designed for accurately measuring the compatibility among possible crowd partitions.

Following the definition of correlation clustering in Eq. (1.5) and its parametrization introduced in Eq. (1.6), the compatibility of an input-output pair of Eq. (1.1), and thus the feature map $\Psi(\mathbf{x}, \mathbf{y})$ is directly described as:

$$F(\mathbf{x}, \mathbf{y}; \mathbf{w}) = \mathbf{w}^T\Psi(\mathbf{x}, \mathbf{y}) = \mathbf{w}^T \sum_{y \in \mathbf{y}} \sum_{a \neq b \in y} [\mathbf{1} - \mathbf{d}(a,b), -\mathbf{d}(a,b)]. \tag{1.7}$$

### Inference and Max Oracle

Solving Correlation Clustering exactly is known to be NP-hard and the problem is also hard to approximate [39]. To deal with this complexity, we adopt a standard greedy procedure [40] where, initially, all pedestrians have their own separate cluster and then, iteratively, the two cluster with the highest correlation are merged. The procedure stops when the best merge would decrease the overall correlation. A similar procedure can be devised for the loss augmented problem of Eq. (1.3) where, at each iteration, the two clusters to be merged are chosen according to both the correlation gain and the score of the loss function. Of course by following a greedy procedure, there is no guarantee to select the most violated constraint. Interestingly enough, Lacoste-Julien *et al.* [37] show that all convergence results known for exact maximizer of the loss augmented problem also hold for approximate maximizers by allowing the

algorithm to iterate longer. For further details, please refer to their work.

### Loss Function

Due to efficiency constraints, the loss function $\Delta(\mathbf{y}_i, \mathbf{y})$ is usually required to decompose with respect to the output. Nevertheless, the iterative nature of aforementioned procedure allows for a more complex and problem-tailored choice. The proposed loss function is based on the *MITRE score* [41]. This score is founded on the understanding that connected components are sufficient to describe groups, inducing a linear amount of positive links (a tree that connects people belonging to the same group) and negative links (a tree that connects groups' connected components). Nevertheless, by working on links, the MITRE score fails to evaluate errors proportionally to the size of the crowd as the number of singletons varies (since singletons have no links). To further square the loss function to the group detection problem, for each pedestrian we add a fake counterpart to which only singletons are connected. Through this shrewdness we can now take into consideration singletons as well when computing the discrepancy between two solutions.

More formally, consider two clustering solutions $\mathbf{y}_i$, $\mathbf{y}$ and a representative of their respective spanning forests $Q$ and $R$. The connected components of $Q$ and $R$ are identified respectively by the set of trees $Q_1, Q_2, \dots$ and $R_1, R_2, \dots$. Note that if the number of elements in $Q_j$ is $|Q_j|$, then only $c(Q_j) \stackrel{\text{def}}{=} |Q_j| - 1$ links are needed in order to create a spanning tree. Let us define $\pi_R(Q_j)$ as the partition of a tree $Q_j$ with respect to the forest $R$, that is the set of subtrees obtained by considering only the membership relations in $Q_j$ also found in $R$. Besides, if $R$ partitions $Q_j$ in $|\pi_R(Q_j)|$ subtrees then $v(Q_j) \stackrel{\text{def}}{=} |\pi_R(Q_j)| - 1$ links are sufficient to restore the original tree. It follows that the recall error for $Q_j$ can be computed as the number of missing links divided by the minimum number of links needed to create that spanning tree. Accounting for all trees $Q_j$ the global recall measure of $\mathbf{y}_i$ is:

$$\mathcal{R}(\mathbf{y}_i) = 1 - \frac{\sum_j v(Q_j)}{\sum_j c(Q_j)} = \frac{\sum_j |Q_j| - |\pi_R(Q_j)|}{\sum_j |Q_j| - 1} \tag{1.8}$$

The precision of $\mathbf{y}_i$ (recall of $\mathbf{y}$) can be computed by exchanging $Q$ and $R$. Given the definition of precision, recall and employing the standard $F$-score $F_1$, the loss is defined as

$$\Delta(\mathbf{y}_i, \mathbf{y}) = 1 - \frac{2\mathcal{R}(\mathbf{y}_i)\mathcal{R}(\mathbf{y})}{\mathcal{R}(\mathbf{y}_i) + \mathcal{R}(\mathbf{y})}. \tag{1.9}$$

## 4.2. Learning to Identify Group Leaders in Crowd

Once groups have been discovered, the crowd structure can be further investigated by discerning its leaders. To underline the key role of leaders in crowd analysis, we quote

a famous case reported in Gustave Le Bon's *The Crowd* [2]:

> During the last strike of the Parisian omnibus employees the arrest of the two leaders directing it was at once sufficient to bring it to an end.

By finding and disconnecting the leaders from the rest of the crowd, efficient containment can be accomplished. At the same time, an influential voice of non-violence in a crowd can lead to a mass sit-in and a strong leader can take control of an emergency situation, initiate movement and guide suitable crowd behavior avoiding panic [42]. Either way, leaders are key subjects to pay attention to when dealing with otherwise unmanageable crowds. Yet, not every crowd qualifies to have a unique influential leader. Many crowds are social gatherings and can be thought of as disconnected groups sharing the same location because of their similar goals, like families in a shopping mall. In such cases, each group has its own leader and it makes sense to restrict the analysis to one group at the time by considering it a *small crowd on its own*.

In this section, we investigate a computational model for the individuation of group leaders in crowded scenes - and of crowd leaders in the case of one group only. We deal with the lack of a formal definition of leadership by learning, in a supervised fashion, a metric space based exclusively on people spatiotemporal information and their partitioning into social groups.

### 4.2.1. Task Formulation

Leader identification amounts at finding the higher scored member for each group, given some leadership scoring criteria. Learning in such a setting might be difficult because evaluating results just by looking at the top scored element makes the objective non-smooth with respect to the input. Nevertheless, it is easy to see that leader identification is lower bounded by leader ranking, where we care about predicting the complete leadership hierarchy. Of course, correctly predicting the leadership hierarchy will also predict the correct leader.

More formally, given as input a social group of size $q$, $\mathbf{x} = \{a, b, \dots\} \in \mathcal{M}$, we want to predict their correct order $\mathbf{y}$ as a permutation of the first $q$ natural numbers, among all possible orderings of that set $\mathcal{Y}(\mathbf{x})$. The leader of $\mathbf{x}$ is the $j$-th member if $\mathbf{y}(j) = 1$. It is easy to see that the best ordering satisfies:

$$\begin{aligned} \mathrm{PR} &= \arg\max_{\mathbf{y}\in\mathcal{Y}(\mathbf{x})} -R(\mathbf{x})^T\mathbf{y} \\ &= \arg\max_{\mathbf{y}\in\mathcal{Y}(\mathbf{x})} -\left((\mathbf{I} - dM)^{-1}\frac{1-d}{|\mathbf{x}|}\mathbf{1}\right)^T\mathbf{y}, \end{aligned} \tag{1.10}$$

where $R$ is a leadership scoring criteria, and –specifically to our case– the PageRank [43]. When the damping factor $d$ is in $(0, 1)$ and $M = (D^{-1}G)^T$ is a column

stochastic matrix, the PageRank has a unique solution. $D$ is the outdegrees diagonal matrix of the graph $G = (\mathbf{x}, f)$. With a slight abuse of notation, $G$ is also the graph matrix such that $G_{a,b} = f(a, b)$ where nodes $(a, b) \in \mathbf{x}$ are the group members and edges $f(a, b)$ encode the probability of $b$ being a leader for $a$ through the generic feature $f$. We choose the PageRank algorithm because of its ability to take advantage of the referential and asymmetric structure of the crowd model [5] to assess the importance of each member.

To let different social aspect intervene in the final ranking, we learn how to combine different contributions from different features. Unluckily, the non-linear nature of PageRank does not allow us to learn how to combine features as edges of a the same graph. Instead we need to compute different PageRanks, each of which works on one graph with edges defined by a single feature, and then aggregate the ranks together. Formally, we generalize $R(\mathbf{x})$ to a linear combination of many PageRanks, each of which is computed on a different feature:

$$R(\mathbf{x}) = \sum_f \mathbf{w}^f \underbrace{\left( (\mathbf{I} - dM^f)^{-1} \frac{1-d}{|\mathbf{x}|} \mathbf{1} \right)}_{R^f(\mathbf{x})\text{: PageRank on feature } f}. \tag{1.11}$$

The features employed in Eq. 1.11 are pairwise time-lagged features leveraging on mutual position, relative speed, DTW. Eventually, member centrality and group size are also considered. Details about the definition of these features can be found in authors' original work [7].

#### *4.2.2. SSVM Adaptation to Leader Identification*

During training, each feature is used to produce a separate ranking of the members of the considered group and Structural SVM is employed to combine different features contribution. In this section we introduce the feature map and the loss function required for this learning to take place.

In order to define the feature map $\Psi(\mathbf{x}, \mathbf{y})$, that is the compatibility of an input-output pair, let $R^*(\mathbf{x}) = [R^{f_1}, R^{f_2}, \dots]$ be the column-wise concatenation different ranks computed on the whole feature set. According to Eq. (1.10) and leveraging on the introduced parameterization, we specify Eq. (1.1) as:

$$F(\mathbf{x}, \mathbf{y}; \mathbf{w}) = \mathbf{w}^T \Psi(\mathbf{x}, \mathbf{y}) = -\mathbf{w}^T R^*(\mathbf{x})^T \mathbf{y}. \tag{1.12}$$

#### Inference and Max Oracle

At test time, for each group, the algorithm returns a ranking of the members, among which the highest will be predicted as the leader. By looking at Eq. (1.10), it is easy to see that the ranking $\mathbf{y}$ maximizing the objective is the one representing the descend-

ing sort of $R(\mathbf{x})$. As a result, inference turns out to be extremely quick even for large groups. To extend the same inference procedure to the computation of the maximization oracle of Eq. (1.3), we need a loss that decomposes with respect to the output elements. Such a loss will be presented in the following paragraph.

### Loss Function

The loss function $\Delta(\mathbf{y}, \mathbf{y}_i)$ should evaluate discrepancies between two predictions. Less obviously, the loss function is also a good place to store and employ prior knowledge. This is because SSVM basically learns how to mimic the loss function by looking merely at inputs (instead of outputs). In our case, the loss function should consider (i) the overall ranking and, even more importantly, (ii) the leader position. One way to handle these requirements is a sum of squared difference of predicted positions, *i.e.* $\Delta(\mathbf{y}, \mathbf{y}_i) = \|\mathbf{y}_i - \mathbf{y}\|^2$. This loss strongly penalizes members whose rank position is predicted further from the true one. With such a loss, all errors are taken into account, but the most costly ones involve members at the boundary of the ranking (firsts and lasts positions) as they can be moved further from their original positions. Thereby, it becomes crucial to be able to correctly predict these positions, while allowing some level of mistakes in the central positions of the hierarchy, where even humans might find the ranking task difficult.

Moreover, the proposed loss is linear w.r.t. the maximization argument $\mathbf{y}$, and a search for the most violating constraint can be accomplished as follows:

$$\begin{aligned} \mathbf{y}_i^* &= \arg\max_{\mathbf{y}\in\mathcal{Y}(\mathbf{x}_i)} \|\mathbf{y}_i - \mathbf{y}\|^2 - \mathbf{w}^T R^*(\mathbf{x})^T \mathbf{y} \\ &= \arg\max_{\mathbf{y}\in\mathcal{Y}(\mathbf{x}_i)} -(2\mathbf{y}_i + R^*(\mathbf{x})\mathbf{w})^T \mathbf{y}, \end{aligned} \tag{1.13}$$

by noting that $\|\mathbf{y}_i\|^2 = \|\mathbf{y}\|^2$ does not depend on the particular choice of $\mathbf{y}$. Trough this shrewdness, the maximization oracle can be efficiently computed as in Eq. (1.12).

## 5. Results on Visual localization of groups and leaders

For the group detection task, we selected two publicly available datasets, namely the *BIWI Walking Pedestrians* dataset [44] and the *Crowds-By-Examples (CBE)* dataset [45]. The former dataset records two low crowded scenes, outside a university and at a bus stop (`eth` and `hotel`). The *CBE* dataset records a medium density crowd outside another university (`student003`, briefly `stu003`) providing different challenges: the density of the pedestrians is significantly high and the presence of multiple entry and exit points. While *BIWI* and *CBE* are standard datasets in crowd analysis, we also use the more recent *Vittorio Emanuele II Gallery (VEIIG)* dataset [46].

Quantitative results of the SSVM approach, introduced in Sec 4.1.2, are given in

| | | hotel | eth | stu003 | GVEII |
|---|---|---|---|---|---|
| $\Delta_{GM}$ | $\mathcal{P}$ | 97.3 | 91.8 | 81.7 | 73.1 |
| | $\mathcal{R}$ | 97.7 | 94.2 | 82.5 | 74.3 |
| $\Delta^{+}_{PW}$ | $\mathcal{P}$ | 89.1 | 91.1 | 82.3 | 70.9 |
| | $\mathcal{R}$ | 91.9 | 83.4 | 74.1 | 71.1 |

(a)

| | SSVM | SVM |
|---|---|---|
| GVEII<br>groups 117 - 75 - 11 | 83.2 | 67.4 |
| stu003<br>groups 87 - 20 - 8 | 82.3 | 78.3 |

(b)

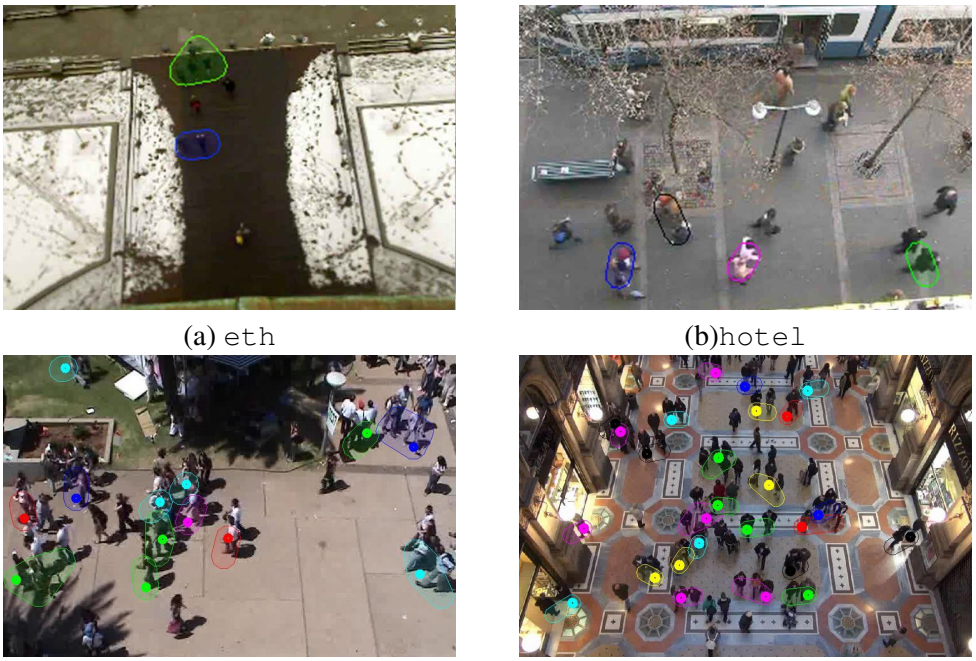
Table 1.1: (a) Quantitative results for visual group detection in terms of precision $\mathcal{P}$ and recall $\mathcal{R}$ computed according to $\Delta_{GM}$ and $\Delta^{+}_{PW}$ respectively. (b) Quantitative accuracy in leader identification.

Tab. 1.1, while visual results are depicted in Figure 1.3. To better appreciate the results of Strucural SVM supervised clustering, we report accuracy in terms of both $G$-MITRE $\Delta_{GM}$ and the more intuitive pairwise loss $\Delta^{+}_{PW}$ [47], which accounts only for positive (intra-group) relations. The slightly lower performances on the `stu003` sequence are due to the higher complexity of the scene and higher density as well.

Leaders have been tested on a subset of the previous datasets, namely `stu003` and `GVEII`. Sociologist manually provided leadership ground truth annotation. Visual examples of the datasets and the achieved results are shown in Figure 1.3. Both `stu003` and `GVEII` present mildly dense but highly group-structured crowds, characterized by the high variability of groups' size and motion patterns. In all scenarios approx the 65% of groups are pairs but triplets and larger groups are present as well. For testing purposes, we employed the ground truth trajectories and group annotation as the input of the SSVM based algorithm of Sec. 4.2. The training of the SSVM is performed independently on every video sequence on the first 20% of the groups. Leader identification accuracy results are reported in terms of binary classification, see Tab. 1.1. We also report an binary SVM baseline, where the leader in a group is the member, among the properly labeled ones, with higher distance from the margin.

## 6. The Predictive Power of Leaders in Social Groups

Until now, we have taken a closing in look at the crowd, going from the whole set of pedestrians to groups and from groups to leaders. In this section we start tackling the reverse problem or, more formally, whether it is possible to recover groups just by knowing their leaders. The final aim of this investigation is a better understanding of the relationship between groups and their leaders, measured as the group detection performance improvement when leaders are used as prior knowledge. Eventually, under the proper investigative tools, we hope to shed some light onto these computationally unexplored questions.



(a) eth (b)hotel

(c) stu003 (d) GVEII

Figure 1.3: Visual results on the employed datasets. Groups are identified by the color of the shape containing their members (first row) and leaders are marked with dots of pertinent colors (bottom row).

## 6.1. Experimental Settings

Going from leaders to groups is a complex task, mainly due to (i) the non-invertible nature of PageRank and (ii) the requirement that groups should be recovered simultaneously and not separately, as in the case of leaders. Considering leaders independently would create inconsistent partitions, dependent on the order followed to visit each leader. As a consequence it is important to approach the problem by considering all leaders as input and by resulting, with global constraints, all respective groups. This requirement, together with the iterative nature of PageRank, makes going from groups to leader a non-invertible path. Figure 1.4 briefly captures the different approaches and the local or global level at which they operate.

Building on the fact that the leaders-to-groups task can be seen as a group detection problem with prior knowledge on the leaders, we propose to computationally tackle it through clustering with seeds. As already stated in Sec. 4.1 Clustering is indeed a proper choice for group detection. By following the same reasoning, we employ both the same features used for groups detection and the same –already learnt– metric

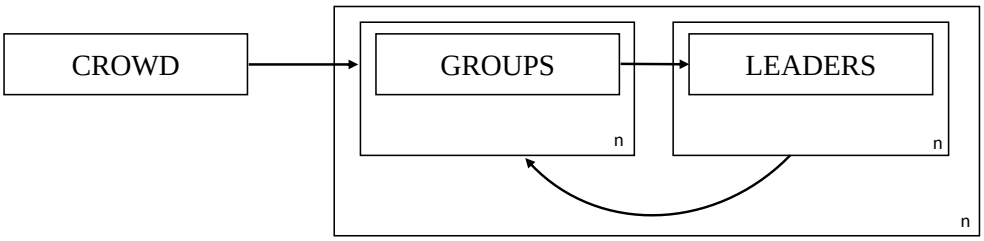



Figure 1.4: Scheme summarizing our different approaches in plate notation. Starting from the crowd, our group detections algorithm retrieves $n$ clusters (*i.e.* groups) simultaneously. Then for each group, and separately from the others, we identify its leader. The last task, recover a group from its leader, can not be accomplished separately for each leader, but requires taking all the leaders and finding all the groups simultaneously.

space. However, Correlation Clustering cannot be adopted because it doesn't support constraining on cluster seeds. In the rest of this section we experiment and validate the k-medoids and the nearest neighbour (NN) clustering algorithms, where leaders can play a role in the initialization step of the clustering. Moreover, as a guarantee for experiments to be sound, we only consider groups with at least three members being from this group size on that leaders can be properly identified.

## 6.2. Leader Centrality in Feature Space

We start by questioning whether a simple k-medoids approach, with cluster medoids initialized on leaders, is able to recover the groups which generated those leaders. The K-medoids partitions members trying to minimize the distance between points labeled to be in a cluster and the medoid point of that cluster. In the case of distances (euclidean or proxemics) on the ground plane, sociologists agree on that the leader doesn't necessarily have to be in the center of its group. We rephrase this question by forcing k-medoids to use our learnt distance in the features space, from Sec. 4.1.

### *6.2.1. Group Recovery Guarantees*

Finding the optimal partition of a multivariate set of data with more than two clusters is known to be NP-hard. In practice, k-medoids is solved fast through Lloyds' algorithm, which iteratively alternates between estimating clusters and their medoids. Under strong hypothesis on the data, applying this iterative algorithm always yields the correct results. Figure 1.5 depicts these hypothesis, better explored below.

Let us initialize a clustering seed for each leader. Since we assume we know the leader, we end up having a leader for each unknown group. Now, define *safe-zone* for each member as the region (in feature space) required by the cluster to avoid containing

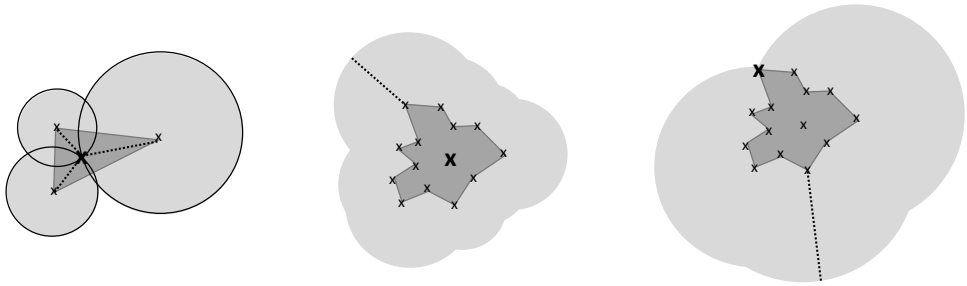


Figure 1.5: K-medoids requirements to guarantee a correct clustering. See text for details.

other initializing seed. This space is a hypersphere in feature space and its radius is determined by the distance of that member from the seed belonging to his group. Moreover, define *safe-zone* for a group the overlap of all safe-zones from all members of that group. Such safe-zones strongly depend on the initialization seeds.

In this setting, the correctness guarantee is data-dependent and not seed-dependent. It is obtained by creating for every group the *worst case safe-zone*, that is a group safe-zone where members contribute through the distances from their furthest member. Note that the worst case safe-zone does not depend from the cluster seed but from members mutual arrangement. Consequently, the correct cluster can be retrieved given any member, not necessarily the leader. Moreover, if the leader is also the medoid of its group, the clustering terminates in one iteration.

### *6.2.2. Validation and Results*

Of course, it rarely happens that we can assure the worst case safe-zone to be empty. In such cases, it is important to study how the group safe-zone changes with respect to the initialization seed, *i.e.* the leader in our setting. As shown in Figure 1.5, a central initialization yields a more homogeneous and averagely smaller group safe-zone. Oppositely, having an initialization on the border of the cluster makes the safe-zone much smaller in some parts as well as much larger in others. Ideally, a central initialization is to be preferred.

In Tab. 1.2 we report results from group detection experiments when leaders were used as initializing seeds for NN and k-medoids. To better understand these results, we correlate them with the concept of leader *centrality*, defined as the one-complement normalized distance between the leader and the group centroid. Intuitively, a more central leader should favor a better group recovery. More formally, centrality is com-

| | | $\mathcal{P}$ | $\mathcal{R}$ | $1-\Delta_{\text{GM}}$ |
|---|---|---|---|---|
| stu003 | CC | 0.8329 | 0.8227 | 0.8272 |
| | NN | 0.8997 | 0.8708 | 0.8848 |
| | k-medoids | 0.8434 | 0.82112 | 0.8319 |
| GVEII | CC | 0.7752 | 0.7932 | 0.7834 |
| | NN | 0.9601 | 0.9528 | 0.9563 |
| | k-medoids | 0.8825 | 0.8690 | 0.8755 |

Table 1.2: Group detection results. Under Nearest Neighbour (NN) and k-medoids, leaders are used as prior information. Conversely, Correlation Clustering reports results of the crowd-to-groups approach presented in Sec. 4.1. Precision and Recall are computed as in Sec. 4.1.2. Having information on leaders always helps in discovering groups, in particular when NN is used for the task.

puted as follow:

$$c(\mathbf{x}_l, \mathbf{x}_c) = 1 - \frac{\|\mathbf{x}_l - \mathbf{x}_c\|}{\max_m \|\mathbf{x}_m - \mathbf{x}_c\|}, \tag{1.14}$$

being $\mathbf{x}_l$ and $\mathbf{x}_c$ the leader and centroid position respectively, and $\mathbf{x}_m$ the position of any other member of that group. Figure 1.6, which depicts group detection correctness, leader centrality and safe-zone violation, confirms that: (i) as long as no safe-zone violation occurs, the leader can always recover its group and (ii) as the safe-zone is violated leader centrality in feature space starts to play a discriminant role. Interestingly, from Tab. 1.2, we observe that having prior information on the leader always brings improvements to group detection performance — particularly in the case of fixed seeds (1 iteration k-medoids, equivalent to NN). While k-medoids can shift the clusters' center, moving *leaders* from their original groups to higher density zones of the crowd, nearest neighbor preserves leader locality, thereby reaching higher scores. This experimental evidence confirms that knowledge about leaders can boost significantly group detection accuracy suggesting these problems should be jointly approached.

## 7. Conclusion

In this chapter we discussed the problem of visually detecting groups and leaders in crowd. We briefly introduced sociological insights to the group and leadership formation process. Moreover, we provided a single structural learning framework for automatically solving these two problems independently. Nevertheless, by experimenting the capability of a trivial clustering algorithm to correctly detect groups (when initialized on leaders), it emerges the strong connection between leaders and groups. The excellent performance of the nearest neighbour clustering seeded on leaders suggests

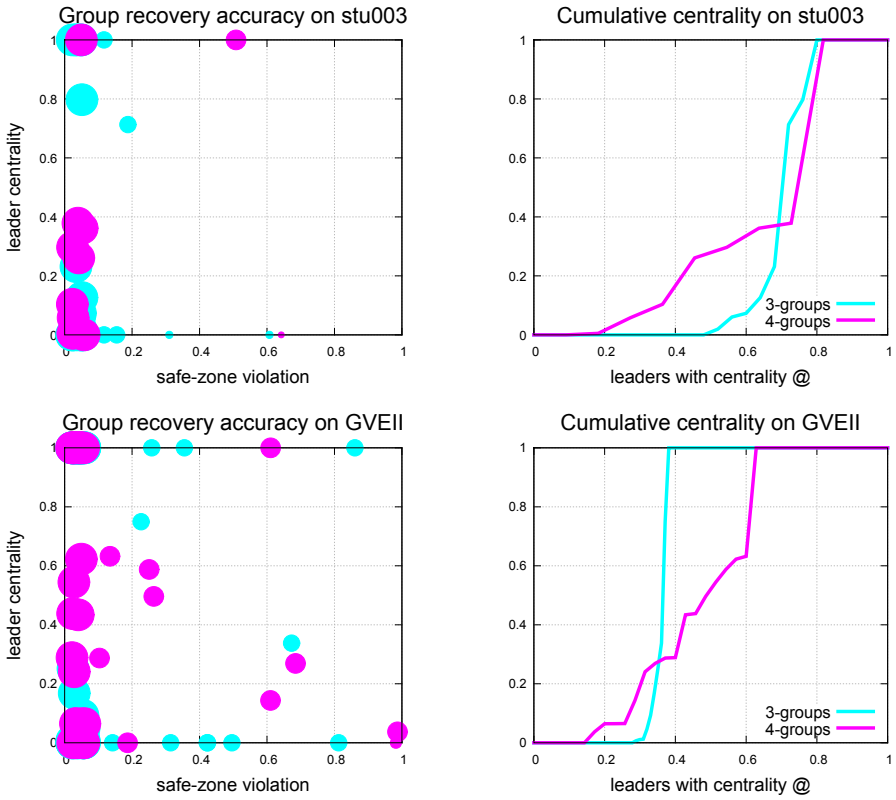



Figure 1.6: Left column: the recall ability of k-medoids is tested on `stu003` and `GVEII` (larger dots mean higher recall). If the safe-zone is not violated, leader centrality doesn't influence the recall. As the violation increases, leader centrality plays an important role. Right column: cumulative distribution of leader with centrality up to a specific value. These figures are obtained by summing points on the scatter plot on the left from bottom to top, always increasing the centrality threshold up to which leaders are considered. Faster growing plots indicate more frequent leader centrality.

that, in the proper feature space, leaders positioning allows for group members to stay closer to their leader than to the other ones. This empirical evidence calls for a joint approach for simultaneously detecting both leaders and groups and we believe this could be a successful future research direction.

## 8. REFERENCE

1. Turner, J.C.; Hogg, M.A.; Oakes, P.J.; Reicher, S.D.; Wetherell, M.S. *Rediscovering the social group: A self-categorization theory*. Basil Blackwell, Cambridge, MA, US **1987**.
2. Le Bon, G. *The Crowd: A Study of the Popular Mind*. Macmillan **1896**.
3. Bandini, S.; Gorrini, A.; Manenti, L.; Vizzari, G. Crowd and Pedestrian Dynamics: Empirical Investigation and Simulation. In *MEASURING BEHAVIOR* **2012**. pp. 308–311.
4. Freud, S. *Massenpsychologie und Ich-Analyse. Die Zukunft einer Illusion*. Psychoanalytischer Verlag, Vienna **1921**.
5. Tarde, G. *The laws of imitation*. Henry Holt and Company, New York, NY, USA **1903**.
6. Solera, F.; Calderara, S.; Cucchiara, R. Socially Constrained Structural Learning for Groups Detection in Crowd. *Transaction on Pattern Analysis and Machine Intelligence* **2015**.
7. Solera, F.; Calderara, S.; Cucchiara, R. Learning to identify leaders in crowd. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* **2015**. pp. 43–48.
8. Moore, B.E.; Ali, S.; Mehran, R.; Shah, M. Visual Crowd Surveillance Through a Hydrodynamics Lens. *Commununications of the ACM* **2011**. *54*, 64–73.
9. Ingham, A.G.; Levinger, G.; Graves, J.; Peckham, V. The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology* **1974**. *10*, 371–384.
10. Helbing, D.; Molnár, P. Social force model for pedestrian dynamics. *Physical Review E* **1995**. *51*, 4282–4286.
11. Moussaid, M.; Perozo, N.; Garnier, S.; Helbing, D.; Theraulaz, G. The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics. *PLoS ONE* **2010**. *5*.
12. Reicher, S.D.; Spears, R.; Postmes, T. A Social Identity Model of Deindividuation Phenomena. *European Review of Social Psychology* **1995**. *6*, 161–198.
13. McPhail, C. *The myth of the madding crowd*. Transaction Publishers **1991**.
14. Manenti, L.; Manzoni, S.; Vizzari, G.; Ohtsuka, K.; Shimura, K. An Agent-Based Proxemic Model for Pedestrian and Group Dynamics: motivations and First Experiments. In *Multi-Agent-Based Simulation XII*. LNCS, Springer Berlin Heidelberg **2012**. pp. 74–89.
15. Calderara, S.; Cucchiara, R. Understanding dyadic interactions applying proxemic theory on videosurveillance trajectories. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)* **2012**. pp. 20–27.
16. Cristani, M.; Paggetti, G.; Vinciarelli, A.; Bazzani, L.; Menegaz, G.; Murino, V. Towards Computational Proxemics: Inferring Social Relations from Interpersonal Distances. In *Proc. IEEE Int'l Conf. Social Computing* **2011**. pp. 290–297.
17. Kendon, A. *Conducting Interaction: patterns of Behavior in Focused Encounters*. Cambridge University Press **1990**.
18. Cristani, M.; Bazzani, L.; Paggetti, G.; Fossati, A.; Tosato, D.; Del Bue, A.; Menegaz, G.; Murino, V. Social interaction discovery by statistical analysis of F-formations. In *Proc. British Machine Vision Conference (BMVC)* **2011**. pp. 1–12.
19. Setti, F.; Russell, C.; Bassetti, C.; Cristani, M. F-formation detection: Individuating free-standing conversational groups in images. *PloS one* **2015**. *10* (5), e0123783.
20. Wang, Y.D.; Wuand, J.K.; Kassim, A.A.; Huang, W.M. Tracking a variable number of human groups in video using probability hypothesis density. In *Proc. Int'l Conf. Pattern Recognition (ICPR)* **2006**. pp. 1127–1130.
21. Feldmann, M.; Fränken, D.; Koch, W. Tracking of extended objects and group targets using random matrices. *IEEE Trans. Signal Processing* **2011**. *59*, 1409–1420.
22. Lin, W.C.; Liu, Y. A lattice-based MRF model for dynamic near-regular texture tracking. *IEEE Trans. Pattern Analysis and Machine Intelligence* **2007**. *29*, 777–792.
23. Pang, S.K.; Li, J.; Godsill, S. Detection and Tracking of Coordinated Groups. *IEEE Trans. Aerospace and Electronic Systems* **2011**. *47*, 472–502.
24. Bazzani, L.; Zanotto, M.; Cristani, M.; Murino, V. Joint Individual-Group Modeling for Tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **2015**. *37* (4), 746–759, doi:10.1109/TPAMI.2014.2353641.
25. Rodriguez, M.; Laptev, I.; Sivic, J.; Audibert, J.Y. Density-aware person detection and tracking in crowds. In *Proc. Int'l Conf. on Computer Vision (ICCV)* **2011**. pp. 2423–2430.

26. Pellegrini, S.; Ess, A.; Van Gool, L. Improving Data Association by Joint Modeling of Pedestrian Trajectories and Groupings. In *Proc. Eur. Conf. Computer Vision (ECCV)* **2010**. pp. 452–465.
27. Yamaguchi, K.; Berg, A.; Ortiz, L.; Berg, T. Who are you with and where are you going? In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR)* **2011**. pp. 1345–1352.
28. Chang, M.C.; Krahnstoever, N.; Ge, W. Probabilistic group-level motion analysis and scenario recognition. In *Proc. Int'l Conf. on Computer Vision (ICCV)* **2011**. pp. 747–754.
29. Andersson, M.; Gudmundsson, J.; Laube, P.; Wolle, T. Reporting leaders and followers among trajectories of moving point objects. *GeoInformatica* **2007**. *12* (4).
30. Yu, T.; Lim, S.N.; Patwardhan, K.; Krahnstoever, N. Monitoring, recognizing and discovering social networks. In *Conference on Computer Vision and Pattern Recognition* **2009**. pp. 1462–1469, doi: 10.1109/CVPR.2009.5206526.
31. Carmi, A.; Mihaylova, L.; Septier, F.; Pang, S.K.; Gurfil, P.; Godsill, S. MCMC-based tracking and identification of leaders in groups. In *International Conference on Computer Vision Workshops* **2011**. pp. 112–119, doi:10.1109/ICCVW.2011.6130232.
32. Kjargaard, M.; Blunck, H.; Wustenberg, M.; Gronbask, K.; Wirz, M.; Roggen, D.; Troster, G. Time-lag method for detecting following and leadership behavior of pedestrians from mobile sensing data. In *International Conference on Pervasive Computing and Communications* **2013**. pp. 56–64, doi: 10.1109/PerCom.2013.6526714.
33. Sanchez-Cortes, D.; Aran, O.; Mast, M.; Gatica-Perez, D. A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups. *IEEE Transactions on Multimedia* **2012**. *14* (3), 816–832, doi:10.1109/TMM.2011.2181941.
34. Tsochantaridis, I.; Joachims, T.; Hofmann, T.; Altun, Y. Large margin methods for structured and interdependent output variables. In *Journal of Machine Learning Research* **2005**. pp. 1453–1484.
35. Joachims, T.; Finley, T.; Yu, C.N.J. Cutting-plane training of structural SVMs. *Machine Learning* **2009**. *77* (1), 27–59.
36. Shalev-Shwartz, S.; Zhang, T. Accelerated proximal stochastic dual coordinate ascent for regularized loss minimization. *Mathematical Programming* **2016**. *155* (1-2), 105–145.
37. Lacoste-Julien, S.; Jaggi, M.; Schmidt, M.; Pletscher, P. Block-coordinate Frank-Wolfe optimization for structural SVMs. *arXiv preprint arXiv:1207.4747* **2012**.
38. Bansal, N.; Blum, A.; Chawla, S. Correlation clustering. *Machine Learning* **2004**. *56* (1-3), 89–113.
39. Tan, J. A note on the inapproximability of correlation clustering. *Information Processing Letters* **2008**.
40. Finley, T.; Joachims, T. Supervised clustering with support vector machines. In *Proceedings of the 22nd international conference on Machine learning* **2005**. ACM, pp. 217–224.
41. Vilain, M.; Burger, J.; Aberdeen, J.; Connolly, D.; Hirschman, L. A model-theoretic coreference scoring scheme. In *Proceedings of the 6th conference on Message understanding* **1995**. Association for Computational Linguistics, pp. 45–52.
42. Challenger, W.; Clegg, W.; Robinson, A. Understanding crowd behaviours: Guidance and lessons identified. *UK Cabinet Office* **2009**.
43. Page, L.; Brin, S.; Motwani, R.; Winograd, T. The PageRank citation ranking: bringing order to the web. *Technical Report. Stanford InfoLab* **1999**.
44. Pellegrini, S.; Ess, A.; Schindler, K.; Van Gool, L. You'll never walk alone: Modeling social behavior for multi-target tracking. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR)* **2009**. pp. 261–268.
45. Lerner, A.; Chrysanthou, Y.; Lischinski, D. Crowds by Example. *Computer Graphics Forum* **2007**. *26*, 655–664.
46. Bandini, S.; Gorrini, A.; Vizzari, G. Towards an integrated approach to crowd analysis and crowd synthesis: a case study and first results. *Pattern Recognition Letters* **2014**. *44*, 16–29.
47. Zanotto, M.; Bazzani, L.; Cristani, M.; Murino, V. Online Bayesian Non-parametrics for Social Group Detection. In *Proc. British Machine Vision Conference (BMVC)* **2012**. pp. 111.1–111.12.